J. Parasitol., 86(5), 2000, p. 960–971

STRAIN TYPING OF *TOXOPLASMA GONDII:* COMPARISON OF ANTIGEN-CODING AND HOUSEKEEPING GENES

T. Lehmann, C. R. Blackston, S. F. Parmley*, J. S. Remington*, and J. P. Dubey†
Division of Parasitic Diseases, CDC, 4770 Buford Higway, MS: F22 Chamblee, Georgia 30341

ABSTRACT: Molecular characterization of *Toxoplasma gondii* isolates is central for understanding differences in disease transmission and manifestations. Only 3 subgroups (lineages) have been discerned with subtle within-lineage variation, permitting low-resolution classification of isolates. Because proteins, coding sequences, and especially antigen-coding genes have been used extensively in previous studies, we focused on sequence variation in introns of housekeeping genes, which may be more informative for phylogenetic analysis because they evolve under lower selection. We compared sequence variation in introns of 5 housekeeping genes with 2 antigen-coding genes. Introns of housekeeping genes were slightly more polymorphic than coding and noncoding regions of antigen-coding genes and only the former showed intralineage variation. Intragenic linkage disequilibrium was complete, but intergenic linkage, although highly significant, was incomplete, suggesting that genes are partially uncoupled. Six of 7 substitutions found within the region coding for the tachyzoite surface antigen, *SAG2*, were nonsynonymous, indicating that diversifying selection acts on this locus. Typing isolates on the basis of housekeeping and antigen-coding genes was consistent, but the phylogenetic relationships among the resulting groups was inconsistent. A cougar isolate typed as lineage II using a restriction fragment length polymorphism assay possessed multiple unique polymorphisms, suggesting that it represents a new lineage. We concluded that introns of housekeeping genes are preferred markers for phylogenetic study, and that multilocus genotyping is preferred for typing parasites, especially from feral or unstudied environments.

Toxoplasmosis is a zoonotic disease caused by the protozoan *Toxoplasma gondii.* Whereas most infections are subclinical, infection during pregnancy may lead to a severe, even fatal congenitally transmitted case. Serious disease also threatens immunocompromised people. Exposure to *T. gondii* is high, as prevalence in adults is typically over 20% (Dubey and Beattie, 1988). Toxoplasmosis is also a serious problem for animal husbandry and is a major cause of abortion in sheep and goats.

T. gondii has been isolated from most mammal and bird species and its range covers virtually the whole world (Dubey and Beattie, 1988). Infection results from eating infected meat that contains the encysted stage (bradyzoite), or ingesting the environmentally resistant stage (oocyst), usually through contaminated food or water. The domestic cat and wild felids serve as definitive hosts, facilitating production of gametes and sexual reproduction. Millions of oocysts can be shed in the feces of a single infected cat (Dubey and Frenkel, 1976). Oocysts survive harsh environmental conditions for many months and can be transported by coprophagous animals such as cockroaches, flies, and earthworms (Frenkel, 1990). Most infections are benign and chronic stages of the parasite persist as long as the host is alive. Congenital transmission also occurs but its importance in natural transmission is probably low (Dubey and Beattie, 1988). *T. gondii* possesses exceptional means to traverse large distances (e.g., in migratory birds) and endure disasters such as droughts, floods, and fires (e.g., in mobile hosts, underground), which explain its high abundance and vast geographical range.

These ecological attributes would allow one to predict high genetic diversity due to a large effective population size, but surprisingly low genetic diversity has been measured among *T. gondii* isolates. Only 6 of 18 allozyme loci were found to be polymorphic in over 60 isolates from different hosts and geographical regions (Darde et al., 1988; 1992). Only 2 alleles segregated in most of these polymorphic loci. Low polymorphism has been measured by restriction fragment length polymorphism (RFLP) (Sibley and Boothroyd, 1992) and by sequence analysis of the *SAG2* (Parmley et al., 1994a), *SAG1* (Rinder et al., 1995), and ssrDNA (Luton et al., 1995; Homan et al., 1997) genes. Three main subgroups (lineages) were discerned that showed remarkable homogeneity within lineage (Howe and Sibley, 1995). High-resolution classification (strain typing) is a prerequisite to study the association between genetically defined subpopulations (''strains'') and attributes such as disease severity and host specificity, but this prerequisite cannot be met with low genetic variability. Genetic diversity of *T. gondii* may have been underestimated in previous studies because of biased sampling of the gene pool or the genome, or both. The available isolates may represent a subset of the gene pool because of overrepresentation of domestic animals and humans or of parasites preadapted to grow in culture or in mice. Likewise, the genetic markers used in previous studies may overrepresent markers of low polymorphism. For example, variation in allozymes refers only to the fraction of mutations that resulted in amino acid replacements that changed the overall charge of the protein. Also, variation of antigen-coding genes (such as *SAG1*) is possibly reduced by purifying selection mediated by the host immune response.

To enhance the resolution of strain typing, we compared genetic diversity in introns of 5 housekeeping genes, which are expected to evolve under minimal (direct) selection, with coding and noncoding regions of 2 antigen-coding genes. We assessed the degree of linkage disequilibrium within and between genes, and tested for selective neutrality of the antigen-coding genes. Finally, we compared the classification of isolates among the different genes, and between them and the current strain typing method.

MATERIALS AND METHODS

Parasite strains

A total of 16 *T. gondii* isolates, comprised of 12 previously characterized laboratory strains (e.g., in Howe and Sibley, 1995) and 4 new

Received 1 November 1999; revised 23 March 2000; accepted 23 March 2000.
* Department of Immunology & Infectious Diseases, Research Institute, Palo Alto Medical Foundation, Palo Alto, California.
† Parasite Biology and Epidemiology Laboratory, Livestock and Poultry Sciences Institute, Agriculture Research Service, USDA, Beltsville, Maryland 20705-2035.

TABLE I. Primers used for each gene and corresponding PCR annealing temperature (see Fig. 1).

Gene and accession No.	Primer location	Primer position	Sequence	Annealing temperature* (°C)
ACT1 U10429	External	1191–1211	GCT GGA CGC GAC CTT ACC GAG	**59**–61
		1980–1959	TCC AGA AAG GAC GAC GTT GCC G	**59**–61
	Internal	1278–1301	CGC GAC ATC AAG GAG AAG CTT TGC	55
		1825–1803	CGG TGA TGA TGT TGC CAT CGG GC	55
B10 AJ223585	External	556–576	CGA AGA CTG GAA ACA GTT CGG	59
		1589–1569	CGT GTC ATG ATG CAT TGC GGG	59
	Internal	747–768	CGA AGC TCG AGA AAA CTG GCA C	55
		1408–1388	CCC AAG TAA CCA CGT ACT GGC	55
FOL1 L08489	External	6661–6680	CAC GTC TGC AAC CTA AAA CC	45–**51**
		7545–7526	TTC GAT AAA GGA CAC AGG GG	45–**51**
	Internal	6779–6799	TTG TGA ACA TCC TCA ACA AGG	55
		7380–7361	CAC GAC CTC AAA ATC CTC GG	55
TUB1 M20024	External	214–233	GCC CTT TCC TTC TCT TTG CG	45–52
		1121–1140	ACT GTT GCA GAT TTG ACG CC	45–52
	Internal	289–308	GAG AGA GGT TAT CAG CAT CC	55
		971–990	CTG GAT ACC ATG TTC CAG GC	55
TUB2 M20025	External	147–167	TGT CGA CTT TTT TCC TTC TGC	45–52
		971–989	TCA GTG TAG TGA CCC TTG G	45–52
	Internal	245–264	AAA TTG GCG CCA AGT TTT GG	55
		792–811	CAT TGA TTC TCT CCA GCT GC	55
MAG1 U09029	External	178–197	TTT AGT TTG GTC ACT CGC CG	50
		1537–1518	CGA AAC TGA TGT CGA AGT CC	50
	Internal 1	178–197	TTT AGT TTG GTC ACT CGC CG	53
		870–851	GCA AGG GTG ACA AAC AAG CC	53
	Internal 2	869–888	GCC ACA GCA ACC GTA GGA TT	53
		1537–1518	CGA AAC TGA TGT CGA AGT CC	53
SAG2 M33572	External	17–36†	GAA ATG TTT CAG GTT GCT GC	50
		1362–1344†	TCA AAG CGT GCA TTA TCG C	50
	Internal 1	17–36†	GAA ATG TTT CAG GTT GCT GC	53
		675–656	TCA AAC CCA GAT GTG ACA CG	53
	Internal 2	624–643	AAC TCA TTG TTC GCG TTC CG	53
		1362–1344†	TCA AAG CGT GCA TTA TCG C	53

* Annealing temperatures shown in bold were used for most isolates. Temperature was lowered when the abundance of the PCR product was insufficient.
† Sequences from Howe et al. (1997).

isolates, was used (Table II). Laboratory strains were arbitrarily selected to represent the 3 lineages (see above). The following new isolates were selected without prior knowledge of their lineage affiliation: an isolate of a human congenitally acquired case from British Columbia, Canada collected during the 1995 outbreak (ManC17); a cougar isolate from the same region collected the following year (Cougar2); a sea otter isolate (Otter3); and a wallaby isolate (Wallaby1). We named strains after their host species, and in the case of human isolates, added a ''C'' or an ''H'' before the unique identification number to denote a congenitally transmitted case or an HIV case, respectively.

Genetic markers

Seven single-copy genes were chosen for sequencing (Tables I, II, Fig. 1). The group named ''housekeeping genes'' encodes proteins that are unlikely to play a role in protective immunity and were selected on the basis of the presence of an intron at least 400 bp long. It includes actin1 (hereafter called *ACT1,* Dobrowolski et al., 1997), α-tubulin (hereafter called *TUB1,* Nagel and Boothroyd, 1988), β-tubulin (hereafter called *TUB2,* Nagel and Boothroyd, 1988), *B10* (Nockemann et al., 1998), and dihydrofolate reductase–thymidylate synthase (hereafter called *FOL1,* Roos, 1993). *B10* encodes an excreted–secreted protein of unknown function. It is immunogenic, but being concealed rather than being presented on the cell surface, it is unlikely to play a role in protective immunity (Nockemann et al., 1998). However, this interpretation may be wrong, so we reexamined our results and conclusions independent of the grouping of *B10* with the housekeeping genes. *FOL1* was implicated in drug resistance in other parasites, but such selection in *T. gondii* is unlikely because toxoplasmosis is a zoonotic disease and parasites, even if ''selected'' in human hosts, would not be transmitted to any other host. The group named ''antigen-coding genes'' encodes proteins that probably play a role in protective immunity and includes major surface antigens of the tachyzoite and the bradyzoite, *SAG2* (Prince et al., 1990; Parmley et al., 1994a) and *MAG1* (Parmley et al., 1994b), respectively. Coding and noncoding regions of antigen-coding genes were sequenced, whereas introns were targeted in housekeeping genes (Fig. 1). A few inconsistencies were found between our sequences and the published sequence of *ACT1, TUB1,* and *MAG1*. Because these inconsistencies were all in sites that were monomorphic among our isolates, we omitted the published sequence from all analyses. All unique sequences were submitted to GenBank (accession numbers: AF247379–AF247383, AF249690–AF249703, AF251810–AF251815).

DNA extraction and sequencing

T. gondii DNA was extracted from tissues of infected mice according to the FastDNA® kit protocol for animal tissue (BIO 101, Vista) and purified using a Qiaquick™ PCR purification kit (Qiagen, Valencia). The presence of *T. gondii* DNA was detected using diagnostic PCR assays (Franzen et al., 1997).

External and internal sets of primers were designed for each gene, on the basis of the published sequence (Table I, Fig. 1). Standard PCR amplification was performed using external primers. PCR conditions were denaturation at 94 C for 5 min, followed by 35 cycles consisting

TABLE II. Typing of isolates to haplotypes of each gene.

Isolate*	Collection site	RFLP lineage	Housekeeping genes				Antigen-coding genes		
			ACT1	*TUB1*	*TUB2*	*FOL1*	*B10*	*MAG1*	*SAG2*
Goat5 (GT-1)	U.S.A.	I	2	1	1	1	1	1	1
ManH6 (CAST)	U.S.A.	I	1	1	1	1	1	1	1
ManC14 (RH)	U.S.A.	I	2	1	1	1	1	1	1
ManC17	Canada	I	1	1	1	1	1	1	1
Cougar2	Canada	II	4	1	4	3	3	4	4
SeaOtter3	U.S.A.	II	5	2	4	4	2	2	2
ManH8 (FORT)	U.S.A.	II	3	2	2	2	2	2	2
ManH10 (MOY)	U.S.A.	II	3	2	2	2	2	2	2
Sheep16 (PDS, Me49)	U.S.A.	II	3	2	2	2	2	2	2
Rabbit4 (BEVERLY)	U.K	II	3	2	2	2	2	2	2
ManH9 (HART)	U.S.A.	III (II/III)†	3	2	2	2	2	3	3
ManH12 (SOU)	U.S.A.	III (II/III)†	3	2	3	2	2	3	3
Wallaby1	U.S.A.	III	3	2	3	2	2	3	3
Chicken7 (C56)	U.S.A.	III	3	2	3	2	2	3	3
ManH11 (OPE)	U.S.A.	III	3	2	3	2	2	3	3
ManH13 (VEG)	U.S.A.	III	3	2	3	2	2	3	3

* Isolates were named after their host (see text for detail). In parentheses is the name used in previous studies (e.g., Howe and Sibley, 1995).
† Shown to be II/III recombinants by Howe and Sibley (1995).

of 94 C for 30 s, 55 C (or as specified in Table 1) for 30 s, and 72 C for 45 s. The last extension step was at 72 C for 10 min. To ensure high specificity and abundance of the PCR product, subsequent amplification was performed using internal primers and the same PCR protocol, except that the number of PCR cycles was 25. The products of the second PCR were purified using the Promega Wizard® PCR Preps DNA purification system (Madison, Wisconsin), and quantified on a 1.5% agarose gel stained with ethidium bromide. The purified PCR product was directly sequenced (i.e., without cloning) in both directions using the internal primers and the ABI PRISM® BigDye™ terminator cycle sequencing kit (Perkin Elmer Applied Biosystems, Foster City, California). The fragments were purified using Centri-Sep columns (Princeton Separations, Adelphia, New Jersey). The sequencing gel was run on the Perkin Elmer ABI Prism 377 sequencer. Sequences were evaluated using Sequence Navigator V 1.0.1. (Perkin Elmer Applied Biosystems) and aligned using Genetics Computer Group Wisconsin Package™ software (Madison, Wisconsin). Lineage determination (genotyping) was performed using RFLP assays at the *SAG*2 locus according to Howe et al. (1997), and later confirmed by the sequence of this gene. However, all *Sau*3AI digests of lineage III isolates resulted in 2 bands (frequently of the same intensity), 1 of the same size as the undigested product, and the other of the expected size of the digested product. These DNA bands were sequenced from 4 presumably lineage III isolates and from one lineage II isolate. As expected, the single band of the lineage II isolate did not have the complete *Sau*3AI restriction site, whereas all the larger bands of the presumed lineage III isolates had the complete restriction site (and their smaller bands fully corresponded to the expected restricted sequences). We concluded, therefore, that incomplete digestion of the DNA by *Sau*3AI was the cause of this pattern.

Sequence analysis

Measures of polymorphism including the number of segregating sites, nucleotide diversity (π), and haplotype diversity were computed for different functional regions of the genes using MEGA 1.02 (Kumar et al., 1993) and DNASP 3 (Rozas and Rozas, 1999). Gaps and a few sites with unreliable information (coded as missing) were ignored unless otherwise stated. Linkage disequilibrium analysis (between pairs of polymorphic sites) was performed using DNASP. Intergenic linkage disequilibrium was performed by joining all sequences of each isolate across its different genes because the genome is haploid. The output file was further analyzed using SAS language (SAS Institute, 1990) to separate intragenic pairs of sites from intergenic pairs of sites. Gene trees and bootstrap tests were computed in MEGA. Sequences of each isolate were joined as described above to construct trees for multiple genes. Because isolates belong to a single species, the p-distance (the proportion of polymorphic sites between 2 sequences) was used to calculate genetic distance and all substitutions were assigned the same weight (Saitou and Nei, 1987).

RESULTS

All 7 genes were polymorphic (Tables II, III, Fig. 2). In addition to 50 polymorphic sites, a total of 7 gaps (insertions/deletions) of sizes between 1 and 10 bp was found across the 7 genes, including 1 in a microsatellite consisting of 7–8 repeats of GT in the *TUB2* gene (Costa et al., 1997). Two bases segregated in all polymorphic sites, except in 1 site of the *B10*, which was also a part of a 10-bp gap. The variability measured here may be higher than in a random collection of *T. gondii* isolates because 12 isolates were selected to represent the 3 lineages. Given that bias, the overall level of polymorphism (average $\pi = 0.0037$, Table III) is rather low. It is possible that a more extensive geographical sampling (all but 1 of our isolates were collected from North America, Table II) would increase the genetic diversity. However, previous studies showed no relation between geography and genetic diversity, despite having isolates from 3 continents (Howe and Sibley, 1995).

The level of polymorphism across the whole gene varied over 6-fold among individual genes, but in contrast to expectations, the difference between housekeeping genes ($\pi = 0.0036$–0.0040) and antigen-coding genes ($\pi = 0.0033$) was minimal (Table III). Introns of the housekeeping genes ($\pi = 0.0037$–0.0041) appeared to be only slightly more polymorphic than noncoding regions of the antigen-coding genes ($\pi = 0.0030$). Thus, despite significant differences in diversity among individual genes (such as between *TUB1* and *TUB2*, Fig. 2), the variation is not structured by the grouping of the genes into housekeeping and antigen-coding groups. This pattern is independent of the grouping of the *B10* gene (Table III). In antigen-coding genes, levels of polymorphism of coding and noncoding regions were also similar, but the majority of substitutions in

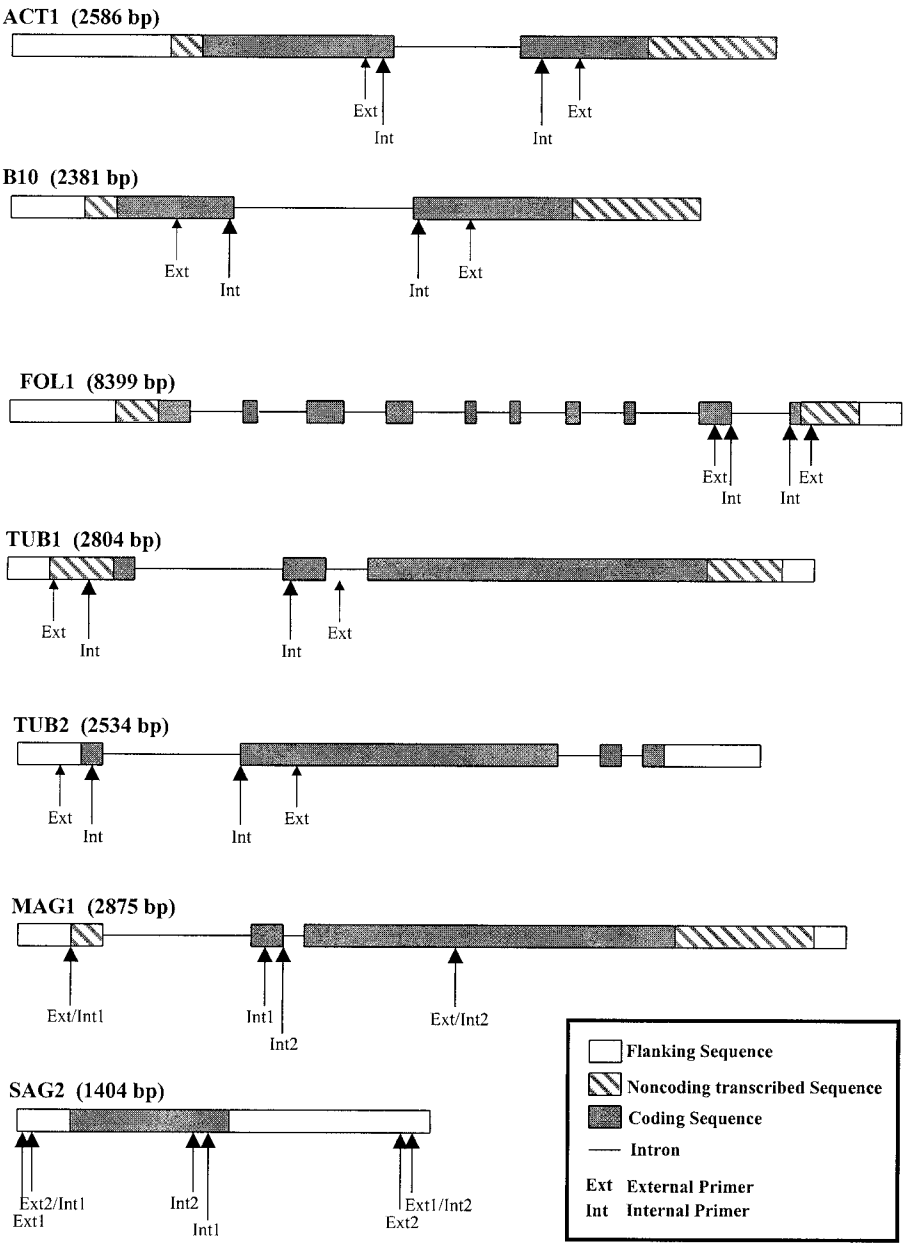


FIGURE 1. Diagrams showing organization of each gene and placement of our primers. Sequences are not drawn to scale.

coding regions of these genes were nonsynonymous (2 of 3 substitutions of *MAG1* and 6 of 8 substitutions of *SAG2* were nonsynonymous).

Linkage disequilibrium within and between genes

Linkage disequilibrium between sites within genes was complete ($|D'| = 1$) and highly significant in all genes including the *MAG1* and the *SAG2*, which comprised up to 1.2 Kb. However, the mean $|D'|$ for sites from different genes ($|D'| = 0.92$) was lower ($P < 0.0001$, Wilcoxon 2-sample test) than 1, indicating that intergenic linkage is incomplete. Linkage disequilibria varied between gene pairs, as no significant D value was found between *SAG2* and all other genes except *MAG1* and *TUB2*, and between other gene pairs (Table IV). *TUB1* and *TUB2* were mapped to chromosome IX, *SAG2* was mapped to chromosome VIII (Sibley et al., 1996), and *ACT1* was mapped to chromosome IB (David Sibley, unpublished data), so physical linkage between some of the genes cannot be ruled out, but it is unlikely given the arbitrary selection of genes and the division of the genome among 11 chromosomes of similar length [2 to 10 Mb, Sibley et al., (1996)]. Tight linkage indicates that recombination has occurred rarely among these isolates and their recent predecessors. Nevertheless, this rate is not negligible, as it was sufficient to uncouple, at least partly, sites from several gene pairs (see more below).

Selective neutrality of antigen-coding genes

Selection that favors diversity (positive selection) implies that the rate of replacement mutations is higher than the rate of silent mutations. Only 3 polymorphic sites were located in the coding regions of *MAG1*, a level of polymorphism that is too low to substantiate a statistical test. Although 2 were replacement substitutions, the rate of synonymous mutations is apparently higher (Table III). In *SAG2*, however, there were 6 replacement mutations and a 3-bp gap that also resulted in an amino acid change, against 2 synonymous mutations (Table III). The nucleotide diversity of nonsynonymous sites (0.0053, without the gap) ap-

TABLE III. Polymorphism in different functional regions of *Toxoplasma gondii* genes.

	Noncoding				Coding synonymous				Coding nonsynonymous				Silent				Total				
Gene	P*	S†/L‡	π§	gap‖	P*	S†/L‡	π§	gap‖	P*	S†/L‡	π§	gap‖	P*	S†/L‡	π§	gap‖	P*	S†/L‡	π§	H#/#¶	gap‖
ACT1	16	7/432	51	0	0	0/9	0	0	0	0/38	0	0	16	7/441	50	0	9	7/479	46	61/5	0
TUB1	2	1/501	9	0	0	0/2	0	0	0	0/8	0	0	2	1/503	9	0	2	1/511	9	46/2	0
TUB2	11	5/463	53	1§§	40	1/14	304	0	0	0/44	0	0	13	6/477	60	1§§	12	6/521	54	78/4	1§§
FOL1	9	5/531	34	2	0	0/4	0	0	0	0/21	0	0	9	5/535	34	2	9	5/556	33	58/4	2
B10	14	8/561	58	1	—		—	—	—		—	—	14	8/561	58	1	14	8/561	58	49/3	1
HK** mean	11	26/2488	41	0.8	19	1/29	61	0	0	0/111	0	0	11	27/2517	42	0.8	10	27/2628	40	58/4	0.8
HK†† mean	9	18/1927	37	0.8	19	1/29	61	0	0	0/111	0	0	10	19/1956	38	0.8	9	19/2067	36	61/4	0.8
MAG1	6	4/708	21	2	7	1/141	33	0	5	2/415	23	0	6	5/849	23	2	6	7/1271	23	74/4	2
SAG2	11	8/746	38	0	14	2/148	39	0	15	6/410	53	1	11	10/894	38	0	12	16/1310	42	74/4	1
AC‡‡ mean	8	12/1454	30	1	10	3/289	36	0	10	8/825	38	0.5	9	15/1743	31	1	9	23/2581	33	74/4	1.5

* Number of polymorphic (segregating) sites per kilobase (based on S and L).
† Number of polymorphic (segregating) sites.
‡ Total sequence length. Estimates of synonymous and nonsynonymous sites were calculated according to Nei and Gojobori (1986) and rounded. Small differences between total sequence length and the sum over the functional regions resulted from exclusion of the whole codon if any sequence had a gap or a questionable nucleotide in that codon, even when the rest of the codon was known, and thus is included in the total gene sequence.
§ Nucleotide diversity (expected heterozygosity per site), $\times 10^{-4}$. Calculated on the basis of Nei (1987), equation 10.5.
‖ Number of gaps (regardless of gap lengths).
Haplotype diversity (expected haplotype heterozygosity), %. Calculated based on Nei (1987), equation 8.4.
¶ Number of unique haplotypes.
** Mean values of housekeeping genes (total values of S and L), the *B10* gene included.
†† Mean values of housekeeping genes (total values of S and L), the *B10* gene excluded.
‡‡ Mean values of antigen-coding genes (total values of S and L).
§§ A gap in a microsatellite, i.e., short-tandem repeat array (see text for details).

Haplotype (n)	*ACT1*	*TUB1*	*TUB2*	*FOL1*	*B10*	*SAG2*	*MAG1*
	1111111			67777	11111	1111	111
	4445566	6	233445	80002	889901122	122234447990122	2499022
	1553513	1	818228	31483	260000616	7146815771145157	9407814
	4361324	8	883033	73770	112304641	6650099098323155	0150224
I (2)	TCTAAAA	A	ACCCGT	GAAAG	GGGCGTCCG	TTATGTCCGAAGCCCT	TTACCGG
I (2)	.G.....	.	.....	.....	.....	.....	.....
II(4)	A..GGG.	G	CGGGCC	CC.GA	A.TG-ATAC	.....GAGCGC..TGC	CC.G.TA
II(1)	A.CGGGT	G	C.GGCC	.C.GA	A.TG-ATAC	.....GAGCGC..TGC	CC.G.TA
II(1)	...G.G.	.	C.GGCC	.CGGA	.A..CA...	.GCAA..GC.CTG..C	..C....
III(5)	A..GGG.	G	C....C	CC.GA	A.TG-ATAC	C.....	....G..
III(1)	A..GGG.	G	CGGGCC	CC.GA	A.TG-ATAC	C.....	....G..
Silent/Replacement	SSSSSSS	S	SSSSSS	SSSSS	SSSSSSSSS	SSRRSRRRRSSSSSSS	SSSSRSR

FIGURE 2. Polymorphic sites in unique haplotypes of each gene. Dots represent identity with the first sequence. The position number refers to the homologous sites in the published sequences (Table 1), despite gaps that may have shifted the location of certain regions in a given sequence. Gaps are not shown except one position, denoted by ''-'' (part of 10-bp gap) in *B10*, which was polymorphic in other isolates.

peared higher than that of synonymous sites at the coding region (0.0039) and that of the flanking regions (0.0038, Table III). The coding region contains 27 amino acids of the signal peptide and 14 amino acids that are clipped as the protein is anchored in the cell membrane (Prince et al., 1990). One replacement substitution was located in the signal peptide, and 1 synonymous substitution was located in the clipped anchored segment. Because those segments are not part of the antigen present on the surface of the parasite, their relevance for positive selection is minimal. The test applied to the remaining 147 codons showed that nucleotide diversity of nonsynonymous sites (0.0064) was significantly higher ($t = 1.66$, 1-tailed test, $P < 0.05$) than that of synonymous sites (0.0011). The difference would be even higher if the gap resulting in amino acid changes was included in the calculations. Because in most proteins the rate of replacement mutations is lower than that of silent mutations (Kreitman and Akashi, 1995), this is considered a conservative test. Lower polymorphism of synonymous sites may also be caused by a strong codon bias or mutation bias toward uneven nucleotide composition, for example, low GC content. The codon bias in the genome of *T. gondii* is minimal and the *SAG2* has the lowest possible bias (Nc = 61, Ellis et al., 1994). The GC content across the genome of *T. gondii* is even (50–55%), and that of *SAG2* is 55.5% (Ellis et al., 1994). Therefore, reduced polymorphism of synonymous sites at *SAG2* by codon and mutation bias is unlikely.

Consistency of typing and phylogenetic patterns based on different genes

Typing of the isolates based on the RFLP assays of Howe et al. (1997) was compared with typing based on sequence variation (Table II). The number of haplotypes per gene (Table II) varied from 2 (*TUB1*) to 5 (*ACT1*). All genes distinguished members of lineage I from the other isolates, whereas only 3 (*TUB2, MAG1,* and *SAG2*) showed differences between lineages II and III. Multilocus ''haplotypes'' comprised 7 groups, of which 6 had unique mutations and the other (ManH9) had no unique mutation, but a unique combination of mutations (a ''mixed'' or hybrid strain). High congruence was found between isolate typing using the RFLP assay and the sequence data (Table II; Fig. 3), except for 1 of the 4 new isolates (see below).

Gene trees based on the proportion of substitutions (p) were constructed using the neighbor joining algorithm (Saitou and Nei, 1987; Kumar et al., 1993) and compared with parsimony networks (Avise, 1994: 122–124), methods that may not produce the same results. Two main clusters of isolates were shown by the parsimony networks of most genes (*TUB1, B10, FOL1,* and *MAG1),* whereas *SAG2* showed three clusters (Fig. 4). Two clades were also shown by all gene trees except *SAG2,* which showed a third clade represented by the cougar isolate (not shown). The networks of *ACT1* and *TUB2* suggested 3 or 4 clusters of isolates separated by nearly equal and relatively small numbers of substitutions. Notably, the two main lineages discerned by *ACT1, TUB2, B10,* and *FOL1* were lineage I versus II/III, whereas the two main lineages discerned by *MAG1* and *SAG2* were lineage I/III versus II. Thus, (a) only 2 lineages are supported by most single gene trees, instead of 3, and (b) the phylogeny is gene dependent, as 2 contradictory phylogenies (with respect to the placement of lineage III) were found. Bootstrap tests of single-locus trees, of the mutational distances between the lineages (not shown), and of the combined 5 housekeeping genes (Fig. 5) indicated conflicting topologies ($P < 0.001$). These results were unaffected by excluding *B10* from the housekeeping genes (not shown).

The cougar isolate, which was typed by the RFLP assay as lineage II, did not cluster with that lineage or with the other lineages (Figs. 4, 5). The number of substitutions separating the cougar isolate from the nearest lineage was higher than that separating at least 1 pair of lineages in 5 of the 7 genes, suggesting that it represents a new lineage (see more below). High bootstrap probabilities (>98%) supported its distinctness (Fig. 5).

TABLE IV. Linkage disequilibrium between pairs of genes expressed as the number of significant tests of disequilibrium over the total number of tests performed. Each test refers to a particular pair of polymorphic sites (one site from each gene). Sets of tests showing no evidence for linkage disequilibrium are in boldface type.

	ACT1	*TUB1*	*TUB2*	*FOL1*	*B10*	*MAG1*
TUB1	4/5					
TUB2	8/30	2/6				
FOL1	16/20	4/4	8/24			
B10	28/35	7/7	14/42	28/28		
MAG1	**0/30**	**0/6**	20/36	**1/24**	**0/42**	
SAG2	**0/50**	**0/10**	36/60	**1/40**	**0/70**	50/60

DISCUSSION

Despite extremely high population abundance and vast geographical range, low genetic diversity has been measured among *T. gondii* isolates (Darde et al., 1988; 1992, Sibley et al., 1996). This diversity is lower than that of other protozoan parasites such as *Giardia duodenalis* (Meloni et al., 1995), *Try-*

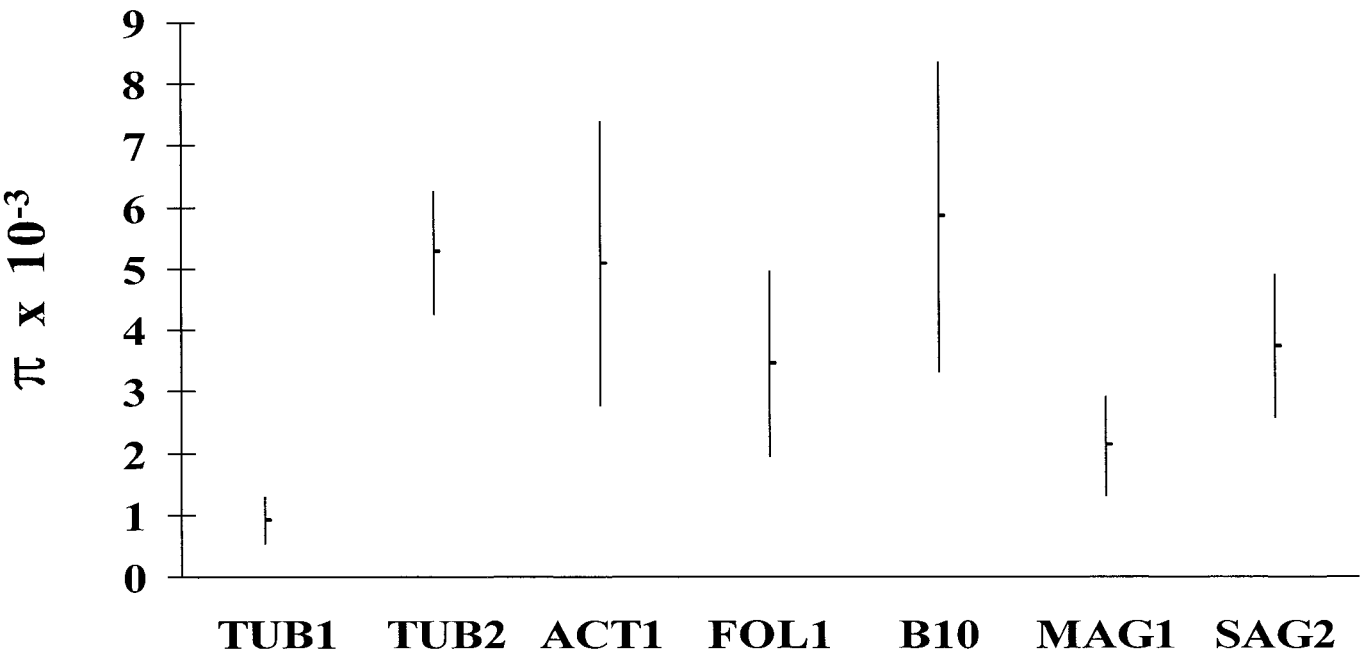


FIGURE 3. Genetic diversity and 95% confidence intervals in introns of each gene (*SAG2* lacks an intron so its flanking regions are used instead).

panosoma cruzi (Marquez et al., 1998), *Tr. brucei* (Penchenier, 1997), *Plasmodium falciparum* (Abderrazak et al., 1999), and *Cryptosporidium* (Morgan et al., 1998), despite far greater host and geographic ranges of *T. gondii.* We addressed the hypothesis that introns of housekeeping genes are more polymorphic than markers previously used and hoped that these introns would serve as informative markers for high-resolution typing of *T. gondii.* High polymorphism and negligible selection are the most important criteria of useful markers.

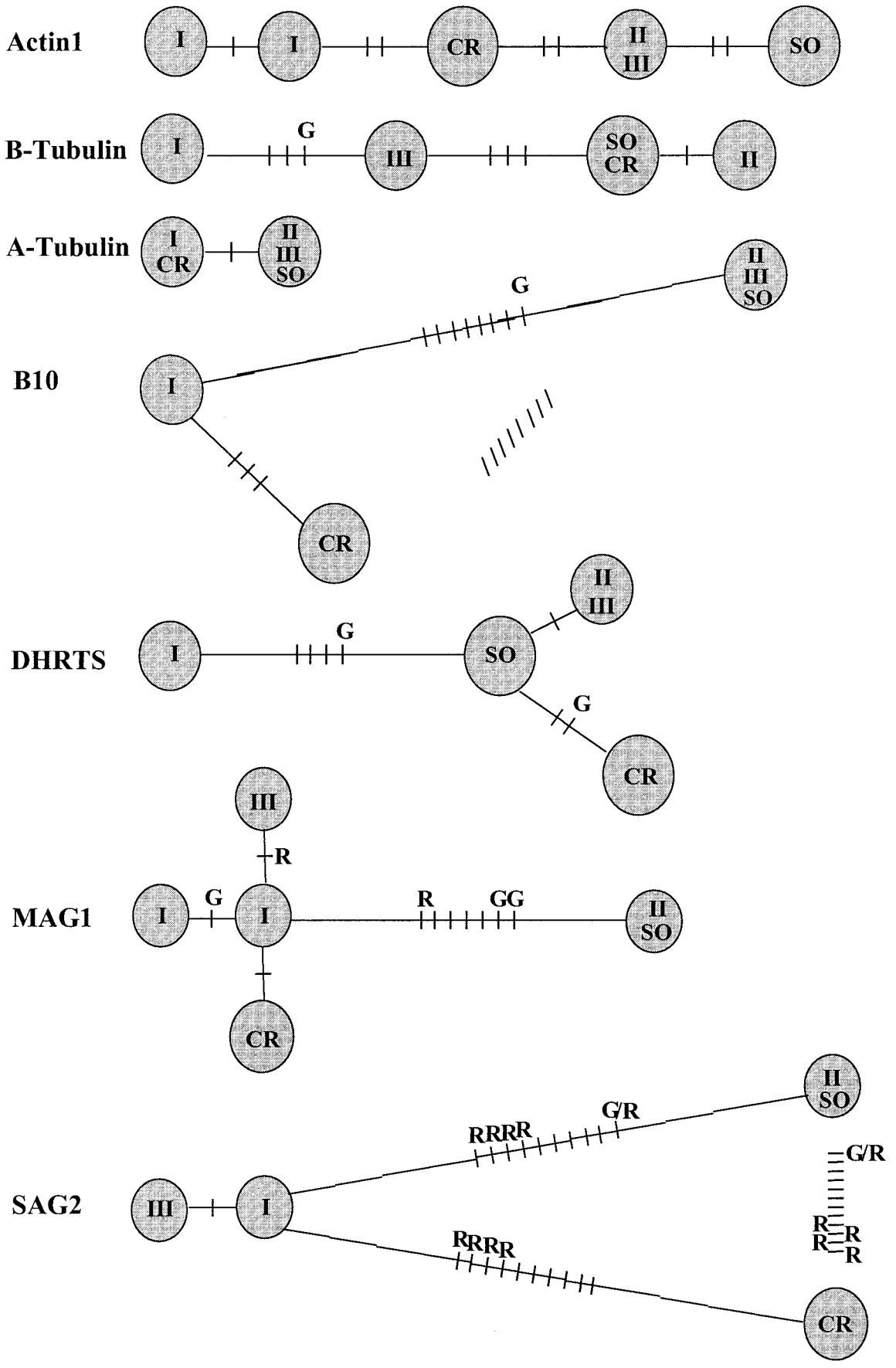
Haplotype diversity (expected allele heterozygosity) varied between 78% (*TUB2*) to 46% (*TUB1*). Only *TUB2, SAG2,* and *MAG1* discerned all 3 lineages from each other. The higher haplotype diversity of the antigen-coding genes was expected because their sequences were 2–3 times longer than those of the housekeeping genes. Surprisingly, no within-lineage variation was detected by the antigen-coding genes (barring the cougar isolate, see below), yet it was detected by the shorter housekeeping genes (*ACT1, TUB2,* and *FOL1*). Therefore, introns of housekeeping genes were more informative than coding and noncoding regions of antigen-coding genes, but the difference between the markers was rather small.

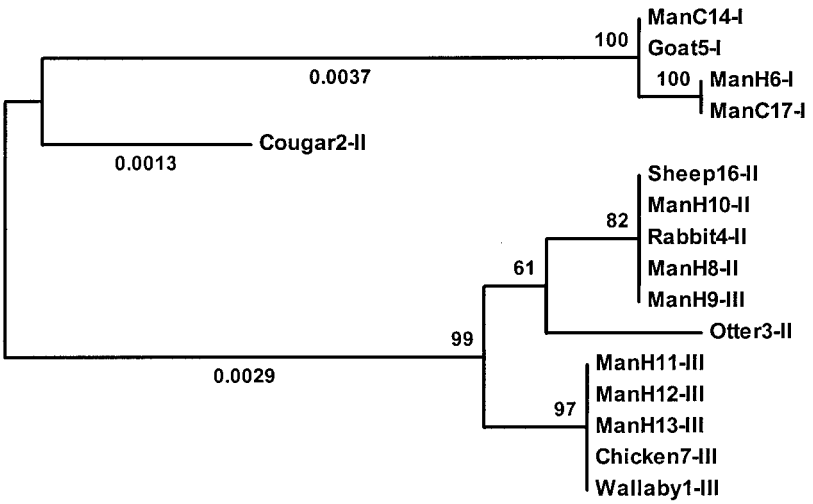
In organisms with strictly clonal population structure, all the genes are completely linked and the effect of selection on any gene is carried throughout the genome. Previous studies reported that ''hybrid stains'' are rare, for example, 5 of 106 (Howe and Sibley, 1995), and concluded that *T. gondii* has an essentially clonal population structure. High and significant linkage disequilibrium supported that view, yet linkage disequilibrium decreased with distance between sites and no significant linkage disequilibrium was found in 8 pairs of genes, suggesting that the recombination rate is not negligible. In addition, isolates were collected many years apart from distant sites, different environments, and different host species. Hence, even if mating within a reproductive unit is random, expected linkage disequilibrium in such a heterogenic sample is higher than nil because of the Wahlund effect (Nei, 1987). These findings and considerations suggest that the effect of sexual reproduction in *T. gondii* cannot be dismissed.

We hypothesized that purifying selection reduces polymorphism in antigen-coding genes. Although limited to only 2 antigen-coding genes, our results showed minimal difference in diversity between these and the introns of housekeeping genes. Moreover, higher polymorphism in nonsynonymous sites than in synonymous sites at *SAG2* indicated that selection favoring antigenic diversity operates on that locus. A higher rate of nonsynonymous mutations (8 replacement vs. 2 silent) was also found in the coding region of *SAG1* (Bulow and Boothroyd, 1991). This mode of selection has been demonstrated for several antigen-coding genes of *P. falciparum* and other parasites (Hughes and Hughes, 1995; Endo et al., 1996; Escalante et al., 1998). This selection is thought to reflect an advantage for a parasite that evades the host immune response despite a previous infection(s) with other variant(s) because of its different antigen(s). Sera of immunized hosts reacted differently to par-

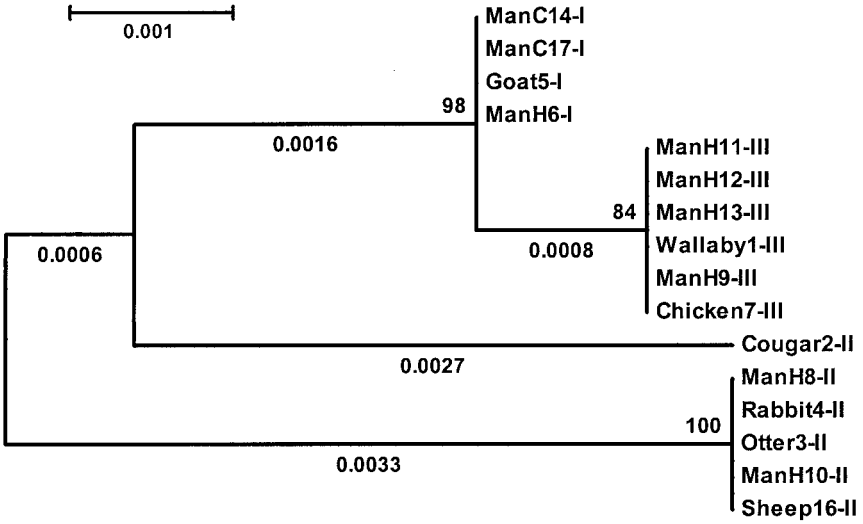
→

FIGURE 4. Parsimony networks based on single genes. All unique haplotypes are marked in circles (lineages I, II, and III by roman numerals, cougar and sea otter isolates by CR and SO, respectively). Short crossing lines represent silent mutations except those marked by R, representing replacements and G, representing gaps. Alternative connections are marked by short crossing lines (representing the number of mutations) without connecting lines. Distance is approximately proportional to genetic distance (assuming equal weight to all mutations) except for alternative connections. The networks do not suppose directionality or origin.

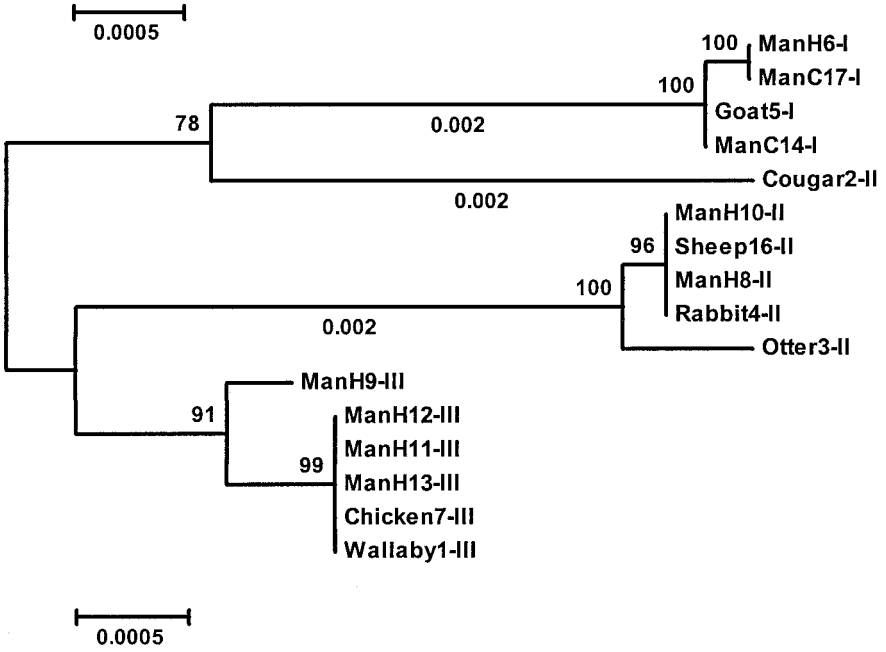

Actin1
I
I
CR
II
III
SO
B-Tubulin
I
G
III
SO
CR
II
A-Tubulin
I
CR
II
III
SO
B10
I
G
II
III
SO
CR
DHRTS
I
G
SO
II
III
G
CR
MAG1
III
R
I
G
I
R
GG
II
SO
CR
SAG2
III
I
RRRR
G/R
II
SO
G/R
RRRR
CR



Panel A



Panel B



Panel C

asites that express different alleles of *SAG2* (Parmley et al., 1994). This explanation implies that recurrent infections are common, and that at least certain hosts carry multiple *T. gondii* variants. Experimentally, mice chronically infected with 1 strain can be infected with a different strain (Araujo et al., 1997), and observations have suggested that reinfection occurs in people (e.g., Coutinho et al., 1982); however, reinfection appears to be uncommon (Dubey and Beattie, 1988; Howe et al., 1997). Alternatively, different antigenic variants may be better adapted to exploit different intermediate hosts or they may represent alternative antigens, each ''equally'' adapted for the same conditions. Notably, selection favoring antigenic diversity in *SAG2* did not result in a larger number of haplotypes, but merely in a greater number of mutations separating lineages I (and III) from II, and these 2 from the cougar isolate. This pattern fits the second explanation better. Other tests of selection, such as Tajima's test, could not be used because 12 of our isolates did not represent a random sample. Regardless of the exact mode of selection, it evidently operates on certain antigen-coding genes of *T. gondii,* and hence it is likely to confound phylogenetic patterns derived from such genes.

Three lineages have been discerned on the basis of RFLP patterns of 5 genes (combined) by Howe and Sibley (1995), in agreement with multilocus allozyme classification (Darde et al., 1988; 1992). Pooling our data across the 7 loci resulted in a similar phylogeny (Fig. 5b, ignoring the lineage formed by the new cougar isolate). This phylogeny, however, was inconsistent with those produced by single genes and by the combined housekeeping and antigen-coding genes (Figs. 4 and 5a, 5b). Thus, lineage III was inseparable from lineage II by *ACT1, TUB1, FOL1,* and *B10,* and both were distinctly separated from lineage I. However, lineage III was inseparable from lineage I by *MAG1* and *SAG2* and both were distinctly separated from lineage II. The existence of 2 lineages instead of 3 and the ''switch'' in affiliation of lineage III agrees with previous results showing 2 alleles of *SAG1* separating lineage I from II/III and 2 alleles at both *SAG2* (closely linked to *SAG1*) and *850,* separating II from I/III (Bulow and Boothroyd, 1991; Parmley et al., 1994; but see Howe and Sibley, 1995). If the tree (Fig. 5c) is true, then the discrepancy between the gene trees may be due to selection on *MAG1* and *SAG2* producing convergence, or slowing down the divergence between lineages I and III. However, if the difference between the groups is adaptive, then it is difficult to explain why most substitutions between them are silent (*MAG1*: 4 of 5, and *SAG2*: 5 of 9) and why lineage II and III are nearly inseparable on the basis of the 5 housekeeping genes. A better explanation is that lineage III represents a recombination between lineage I and II. Thus, *MAG1* and *SAG2* originated from lineage I, whereas the other 5 genes originated from lineage II. The recombination hypothesis accounts for the flaws of the selection hypothesis, but does not explain why the difference is between housekeeping and antigen-coding genes. The situation is further complicated by sequences of the genes *BSR4* and *SAG4* showing lineage I inseparable from lineage II and both distinctly separated from lineage III (M. Grigg, S. Bonnefoy, A. Hehl, and J. Boothroyd, personal communication). These results suggest that lineage III does not represent a single recombination event between lineages I and II, but that there were multiple recombination events between 2 or more diverged haplotypes carriers. Finally, the distinct *SAG2* haplotype of the cougar isolate suggests that there were at least 3 (rather than 2) diverged parental types. Interestingly, the haplotypes at the other genes of this isolate have not diverged as much from those of lineage I (e.g., *MAG1*) or II/III (e.g., *TUB2* or *FOL1*), as depicted in Figure 4. These results stress that recombination cannot be ignored in the phylogeny of *T. gondii.*

All 106 isolates analyzed by Howe and Sibley (1995) clustered in 3 lineages. The new cougar isolate, however, possessed multiple unique polymorphisms and consequently formed a separate clade, which was supported by bootstrap probabilities higher than 95% (Fig. 5). In 5 of the 7 genes, genetic distance between this isolate and its nearest lineage was larger than the distance separating (at least) 1 pair of the 3 recognized lineages. Different from the ''natural recombinants'' or isolates with rare alleles (Darde et al., 1992; Howe and Sibley, 1995) that could have been produced by a recombination or a single mutation, the sequences reveal multiple mutations (in multiple genes) substantiating a large genetic distance between alleles of the cougar isolate and those of other lineages. Genetic distance between alleles cannot be estimated in allozyme data (Darde et al., 1992) and it was not estimated in the RFLP data (Howe and Sibley, 1995); thus, it cannot be determined if these rare alleles are more related to the alleles representing the lineages to which they were assigned or not. Sequence data, however, facilitate accurate assessment of the genetic distance between alleles, suggesting the cougar isolate represents a new lineage. This inference is independent of the frequency of the isolate and of the presence of other closely related isolates. However, if we underestimated intralineage variation by missing other variants of lineage I or II, this variation would increase and the distinction between the lineages could decrease, possibly uniting the cougar isolate with another lineage. Whereas this possibility cannot be ruled out, we consider it unlikely as previous studies did not reveal greater within-lineage variation despite having larger number of isolates (e.g., Howe and Sibley, 1995). Typing the cougar isolate as lineage II highlights a limitation of the sensitive and efficient RFLP assays, which assume that *T. gondii* is composed of only 3 (entirely clonal) lineages. Although probably correct in most human instances, it will misclassify isolates representing new lineages and certain recombinants. A new lineage isolated from a ''wild'' host suggests that genetic diversity of *T. gondii* is higher than estimated because most isolates were collected from humans, and domestic and peridomestic animals, whereas ''wild'' *T. gondii* were underrepresented. If that is the case, it remains to be determined

←

FIGURE 5. Phylogenetic relationships among isolates on the basis of the 5 housekeeping genes (panel A), the 2 antigen-coding genes (panel B), and all 7 genes (panel C). Excluding the *B10* gene resulted in identical topology (not shown). Trees were drawn on the basis of π using the Neighbor Joining algorithm (see Materials and Methods). Bootstrap values (above 50%) and branch lengths (above 0.0001) are shown above and below the branches, respectively.

what keeps the wild isolates from spreading into the domestic and peridomestic environment.

In conclusion, variation in introns of housekeeping genes was not significantly higher than that in coding or noncoding regions of antigen-coding genes. The low levels of genetic diversity measured in previous studies of *T. gondii*, therefore, is not due to markers of low polymorphism. Why a ubiquitous parasite with an exceptionally broad host range possesses so low a genetic diversity remains to be answered. Introns of housekeeping genes are preferred markers for phylogenetic inferences because selection, as demonstrated on *SAG2*, is unlikely to operate on them. Selfing (sexual reproduction involving gametes of the same individual) and clonal reproduction predominate in *T. gondii*, as evident from high linkage disequilibrium, but the effect of sexual reproduction (uniting gametes of different parasites) should not be dismissed, as intergenic linkage is incomplete, and typing an isolate on the basis of different genes can result in different grouping. Finally, genetic diversity of *T. gondii* in ''wild'' hosts may be higher than current estimates, as suggested by the cougar isolate.

ACKNOWLEDGMENTS

We are grateful to Alexander da Silva, Iaci Moura, Norman Pieniazek, Jane Carter, and especially Ellen Dotson (CDC) for helping us with the laboratory work. Ben Beard, Altaf Lal, Vance Dietz, Dennis Juranek, Sue Binder, Tom Navin, and especially Marianna Wilson (CDC) provided valuable advice on various aspects of this project. We thank Fred Simard, Ananias Escalante, Martin Donnelly, Odette Mukabayire (CDC), David Sibley (Washington University), John Boothroyd, and Mike Grigg (Stanford University) for their critical comments, and the latter two for sharing their unpublished data. Allen Hightower and colleagues (CDC) provided excellent computer support. Helpful comments were also provided by an anonymous reviewer.

LITERATURE CITED

ABDERRAZAK, S. B., S. H. QARY, A. A. LAL, M. F. BOSSENO, P. FORCE-BARGE, J. P. DUJARDIN, T. FANDEUR, J. F. MOLEZ, F. KJELLBERG, F. J. AYALA, AND M. TIBAYRENC. 1999. *Plasmodium falciparum*: Population genetic analysis by multilocus enzyme electrophoresis and other molecular markers. Experimental Parasitology **92:** 232–238.

ARAUJO, F., T. SLIFER, AND S. KIM. 1997. Chronic infection with *Toxoplasma gondii* does not prevent acute disease or colonization of the brain with tissue cysts following reinfection with different strains of the parasite. Journal of Parasitology **83:** 521–522.

AVISE, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York, 511 p.

BULOW, R., AND J. C. BOOTHROYD. 1991. Protection of mice from fatal *Toxoplasma gondii* infection by immunization with P30 antigen in liposomes. Journal of Immunology **147:** 3496–3500.

COSTA, J. M., M. L. DARDE, B. ASSOULINE, M. VIDAUD, AND S. BRETAGNE. 1997. Microsatellite in the beta-tubulin gene of *Toxoplasma gondii* as a new genetic marker for use in direct screening of amniotic fluids. Journal of Clinical Microbiology **35:** 2542–2545.

COUTINHO, S. G., M. A. LEITE, M. R. AMENDOEIRA, AND M. C. MARZOCHI. 1982. Concomitant cases of acquired toxoplasmosis in children of a single family: evidence of reinfection. Journal of Infectious Diseases **146:** 30–33.

DARDE, M. L., B. BOUTEILLE, AND M. PESTRE-ALEXANDRE. 1988. Isoenzymic characterization of seven strains of *Toxoplasma gondii* by isoelectrofocusing in polyacrylamide gels. American Journal of Tropical Medicine and Hygiene **39:** 551–558.

———, ———, AND ———. 1992. Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiological implications. Journal of Parasitology **78:** 786–794.

DOBROWOLSKI, J. M., I. R. NIESMAN, AND L. D. SIBLEY. 1997. Actin in the parasite *Toxoplasma gondii* is encoded by a single copy gene, ACT1, and exists primarily in a globular form. Cell Motility and the Cytoskeleton **37:** 253–262.

DUBEY, J. P., AND C. P. BEATTIE. 1988. Toxoplasmosis of animals and man. CRC Press, Boca Raton, Florida, 220 p.

———, AND J. K. FRENKEL. 1976. Feline toxoplasmosis from acutely infected mice and the development of *Toxoplasma* cysts. Journal of Protozoology **23:** 537–546.

———, AND ———. 1998. Toxoplasmosis of rats: A review, with considerations of their value as an animal model and their possible role in epidemiology. Veterinary Parasitology **77:** 1–32.

ELLIS, J. T., D. A. MORRISON, D. AVERY, AND A. M. JOHNSON. 1994. Codon usage and bias among individual genes of the coccidia and piroplasms. Parasitology **109:** 265–272.

ENDO, T., K. IKEO, AND T. GOJOBORI. 1996. Large-scale search for genes on which positive selection may operate. Molecular Biology and Evolution **13:** 685–690.

ESCALANTE, A. A., A. A. LAL, F. J. AYALA. 1998. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. Genetics **149:** 189–202.

FRANZEN, C., M. ALTFELD, P. HEGENER, P. HARTMANN, G. ARENDT, H. JABLONOWSKI, J. ROCKSTROH, V. DIEHL, B. SALZBERGER, AND G. FATKENHEUER. 1997. Limited value of PCR for detection of *Toxoplasma gondii* in blood from human immunodeficiency virus-infected patients. Journal of Clinical Microbiology **35:** 2639–2641.

FRENKEL, J. K. 1990. Transmission of toxoplasmosis and the role of immunity in limiting transmission and illness. Journal of the American Veterinary Medical Association **196:** 233–240.

HOMAN, W. L., L. LIMPER, M. VERLAAN, A. BORST, M. VERCAMMEN, AND F. KNAPEN. 1997. Comparison of the internal transcribed spacer, ITS 1, from *Toxoplasma gondii* isolates and *Neospora caninum*. Parasitology Research **83:** 285–289.

HOWE, D. K., S. HONORE, F. DEROUIN, AND L. D. SIBLEY. 1997. Determination of genotypes of *Toxoplasma gondii* strains isolated from patients with toxoplasmosis. Journal of Clinical Microbiology **35:** 1411–1414.

———, AND L. D. SIBLEY. 1995. *Toxoplasma gondii* comprises three clonal lineages: correlation of parasite genotype with human disease. Journal of Infectious Diseases **172:** 1561–1566.

HUGHES, M. K., AND A. L. HUGHES. 1995. Natural selection on *Plasmodium* surface proteins. Molecular and Biochemical Parasitology **71:** 99–113.

KREITMAN, M., AND H. AKASHI. 1995. Molecular evidence for natural selection. Annual Review of Ecology and Systematics **26:** 403–422.

KUMAR, S., K. TAMURA, AND M. NEI. 1993. MEGA: Molecular evolutionary genetics analysis version 1.01. The Pennsylvania State University, University Park, Pennsylvania, 130 p.

LUTON, K., M. GLEESON, AND A. M. JOHNSON. 1995. rRNA gene sequence heterogeneity among *Toxoplasma gondii* strains. Parasitology Research **81:** 310–315.

MARQUEZ, E., M. ARCOS-BURGOS, O. TRIANA, J. MORENO, AND N. JARAMILLO. 1998. Clonal population structure of Colombian sylvatic *Trypanosoma cruzi*. Journal of Parasitology **84:** 1143–1149.

MELONI, B. P., A. J. LYMBERY, AND R. C. THOMPSON. 1995. Genetic characterization of isolates of *Giardia duodenalis* by enzyme electrophoresis: Implications for reproductive biology, population structure, taxonomy, and epidemiology. Journal of Parasitology **81:** 368–383.

MORGAN, U. M., K. D. SARGENT, P. DEPLAZES, D. A. FORBES, F. SPANO, H. HERTZBERG, A. ELLIOT, AND R. C. THOMPSON. 1998. Molecular characterization of *Cryptosporidium* from various hosts. Parasitology **117:** 31–37.

NAGEL, S. D., AND J. C. BOOTHROYD. 1988. The alpha- and beta-tubulins of *Toxoplasma gondii* are encoded by single copy genes containing multiple introns. Molecular Biochemistry and Parasitology **29:** 261–273.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, New York, 512 p.

NEI, M., AND T. GOJOBORI. 1986. Simple methods for estimating the

numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution **3:** 418–426.

NOCKEMANN, S., H. DLUGONSKA, B. HENRICH, A. KITZEROW, AND W. DAUBENER. 1998. Expression, characterization and serological reactivity of a 41 kDa excreted-secreted antigen (ESA) from *Toxoplasma gondii*. Molecular Biochemistry and Parasitology **97:** 109–121.

PARMLEY, S. F., U. GROSS, A. SUCHARCZUK, T. WINDECK, G. D. SGARLATO, AND J. S. REMINGTON. 1994a. Two alleles of the gene encoding surface antigen P22 in 25 strains of *Toxoplasma gondii*. Journal of Parasitology **80:** 293–301.

———, Y. SHUMIN, G. HARTH, D. L. SIBLEY, A. SUCHARCZUK, AND J. S. REMINGTON. 1994b. Molecular characterization of a 65 kilodalton *Toxoplasma gondii* antigen expressed abundantly in the matrix of tissue cysts. Journal of Parasitology **80:** 293–301.

PENCHENIER, L., F. MATHLEU-DAUDE, C. BRENGUES, A.L. BANULS, AND M. TIBAYRENC. 1997. Population structure of *Trypanosoma brucei* s.l. in Cote d'Ivoire assayed by multilocus enzyme electrophoresis: epidemiological and taxonomical considerations. Journal of Parasitology **83:** 19–22.

PRINCE, J. B., K. L. AUER, J. HUSKINSON, S. F. PARMLEY, F. G. ARAUJO, AND J. S. REMINGTON. 1990. Cloning, expression, and cDNA sequence of surface antigen P22 from *Toxoplasma gondii*. Molecular Biochemistry and Parasitology **43:** 97–106.

RINDER, H., A. THOMSCHKE, M. L. DARDE, AND T. LOSCHER. 1995. Specific DNA polymorphisms discriminate between virulence and nonvirulence to mice in nine *Toxoplasma gondii* strains. Molecular Biochemistry and Parasitology **69:** 123–126.

ROOS, D. S. 1993. Primary structure of the dihydrofolate reductase-thymidylate synthase gene from *Toxoplasma gondii*. Journal of Biological Chemistry **268:** 6269–6280.

ROZAS, J., AND R. ROZAS. 1999. DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

SAITOU N., AND M. NEI. 1987. The neighbor joining method: a new method for constructing phylogenetic trees. Molecular Biology and Evolution **4:** 514–525.

SAS INSTITUTE INC. 1990. SAS Language: References, Version 6. Cary, North Carolina.

SIBLEY, L. D., AND J. C. BOOTHROYD. 1992. Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. Nature **359:** 82–85.

———, D. K. HOWE, K.-L. WAN, S. KHAN, M. A. ASLETT, AND J. W. AJIOKA. 1996. *Toxoplasma* as a model genetic system. *In* Molecular biology of parasitic protozoa, D. F. Smith and M. Parsons (eds). IRL Press, Oxford, U.K., p. 55–74.